# Virtual Assistants

## A Study on the Usability and User Perception of Customer Service Systems for E-Commerce

*Valerie Claessen, Adrian Schmidt, Tamara Heck*

Heinrich-Heine-University Düsseldorf, Germany
{valerie.claessen, adrian.schmidt, tamara.heck}@hhu.de

**Abstract**

Virtual assistants are able to support users in finding the right information. These programs use natural language processing, learning techniques and social abilities to offer adequate usability experiences for users. In e-commerce, virtual assistants are applied to support users in finding appropriate service information or products. This work evaluates the information service quality of three virtual assistants on e-commerce websites. The analyzed aspects cover service quality as well as user perception of virtual assistant systems. First results show that the apparent technology used in the construction of the virtual assistant has a substantial influence on user experience, by the users' perceived interaction with the assistant becoming more intuitive and therefore more enjoyable. Overall, all assistants were met with a general sense of enthusiasm. However, scores on the usefulness of the services show that they need to be improved regarding several relevant features.

**Keywords:** virtual assistant; intelligent agent; user study; human computer interaction; human information behavior

In: M. Gäde/V. Trkulja/V. Petras (Eds.): Everything Changes, Everything Stays the Same? Understanding Information Spaces. Proceedings of the 15th International Symposium of Information Science (ISI 2017), Berlin, 13th—15th March 2017. Glückstadt: Verlag Werner Hülsbusch, pp. 116—130.

# 1 Introduction

Human to human communication is progressively being replaced by human to computer interaction, while the prevalence of online sales and distribution is increasing rapidly. Advances in machine learning and data mining offer new opportunities in e-commerce to support and assist customers. One ongoing trend is the application of virtual assistants. Market research Gartner prognoses immense developments in intelligent apps, which assist companies and their employees (Panetta, 2016). In e-commerce, virtual assistants shall make users' tasks easier as well. Such programs are a further development of search engines (Chao et al., 2016) that have specific features like autonomy, proactiveness and learnability, and "assist and replace users in executing the time-consuming task of compiling information" (ibid.: 118). Those agents use natural language processing to communicate with users. In research literature, different terms are used to describe those agents: Virtual or intelligent agents, assistant or avatars, bots or even virtual personal agents, which might refer to software or real human assistants (compare McGoldrick, Keeling & Beatty 2008 for an overview). Modern programs have a further relevant feature, i.e. social abilities: An "intelligent agent interface [that] imitates several dimensions of social interaction" (Kuligowska & Lasek, 2005: 4). This software is socially interactive and simulates a kind of human to human dialogue. Such assistants are applied in the fields of education and e-learning, e-commerce, administration, and consulting. They shall advice users and/or lead them to buy products or services. In the following work, we use the term virtual assistant (VA) and concentrate on software that uses any natural language processing and has a graphical human-like, i.e. personified character. The evaluated VAs are current examples of assistants by three leading German companies active in the VA software development.[1] They are all inserted on e-commerce websites and support users in finding products or service information. Research focuses on the performance of these VAs and their user perception and usefulness.

1 For an overview of VAs and companies see: https://www.chatbots.org/country/de/.

## 1.1 Related work

Studies on virtual assistants focus on diverse research aspects, like natural language processing and recommender system techniques (Chang, Lee & Wang, 2016; García-Serrano, Martínez & Hernández, 2004), VA interface and its user perception (Von der Pütten et al., 2010) and VA implementation for specific purposes like in e-learning or healthcare services (Ahamed et al., 2006). Compared to "physically-embodied agents" (also called robots), virtual agents are purely digital (Li, 2015). However, many research papers only refer to virtual agents or VAs as the software, i.e. the part of a system being able to process natural language and learning techniques from either text or spoken language. Other popular terms for such technical programs are "chatbot" or "chatterbot". Besides this definition, other researcher assign a morphology to VAs, thus VAs can be either anthropomorphic, zoomorphic, caricatured or functional (Fong, Nourbakhsh & Dautenhahn, 2003). Furthermore, those VAs can have different types of modalities depending on their morphology: For example, an anthropomorphic VA (also called personified agent) can have an entire body, or just arms and a head (Li, 2015). VAs are distinguished from avatars, which represent and reflect the behavior of a human being, like avatars in Second Life (Von der Pütten et al., 2010). In this paper, we refer to VAs, which are represented by a digital anthropomorphic character. Besides research on VA techniques, studies assessing anthropomorphic VAs concentrate on users' perception of these personified agent characters. In their user study, Shiban et al. (2015) tested the influence of two human-like agents (male and female) on e-learners' enjoyment and interest. They detected no great differences between the agents, and students had similar performance even without the presence of any agent. Etemad-Sajadi (2014) studied ten diverse aspects (e.g. trust, ease of use, aesthetic) based on an expanded technology acceptance model (TAM) of a personified VA implemented on a restaurant's website. The author's analysis and his deployment of a user survey had similar intentions as the following study, with a focus on the usefulness and perception of a VAs. Results showed that usefulness and enjoyment positively influenced the users' likelihood of revisiting the restaurant's website. Furthermore, the aesthetic aspect of a VA and a user's enjoyment have a strong interrelation, as well as trust and usefulness. Other studies, which implement and evaluate VAs, also have a focus on e-commerce and marketing aspects (e.g. Keeling, McGoldrick & Beatty, 2010). Their research questions mostly focus on the impact of VAs on the

quality of websites. The proposed work will have a deeper focus on human to VA interaction and a user's information need, leading to questions such as 'how do users evaluate information by a VA and how is information perceived via a VA?'. The study compares the usability of diverse VAs, regarding not only language processing techniques, but VA's performance and its perception from a user perspective. It will give first insights into human to VA interaction and its influence in user information behavior. The first results introduced in this paper concentrate on the perceived quality of the system and perceived quality of the content.

## 1.2 Research question and VAs

The following work addresses the VAs' performance and perception, and answers the following research question: How do VAs perform according to relevant information service criteria?

To answer this question, we evaluated and compared examples of three VAs, based on an empirical user study as well as a heuristic analysis (Sarodnick & Brau, 2011). We searched for companies offering German-speaking VA software technology and chose current and similar VAs deployed on German websites from three leading providers[2]: Carla[3] by Kauz Linguistic Technologies, Clara by Novomind[4], and Jana[5] by Artificial Solutions[6].

Carla has been active since early 2016 and was designed to be a fictional worker of a chocolate manufacturer. With Carla, customers are able to be recommended chocolate according to their preferences. The drawn female VA answers all kinds of questions regarding the creation of the offered goods, the manufacturer, as well as shipping and billing. The 3D rendered female VA of Otto, Clara[7], went online in May 2013. Similar to Carla, Clara[8] was designed to inform users of products and services of the Otto GmbH. Lastly, we considered the VA of the E-Post, Jana, which was deployed in December 2013. Jana is represented by a picture of a real woman. She has

2 For an overview see https://www.chatbots.org/country/de.

3 www.kauz.net/Carla

4 www.novomind.com

5 www.epost.de/privatkunden/hilfe-jana.html

6 www.artificial-solutions.com

7 www.otto.de/kontakt

8 To avoid any confusion with the VA Carla, Clara is called Otto from now on.

the role of an E-Post consultant and answers questions regarding the delivery and receiving of electronic mail by the Deutsche Post DHL Group.

# 2 Methods

The evaluation of the three VAs is based on the information service evaluation (ISE) model (Schumann & Stock, 2014) because it allows for a holistic view of diverse aspects of a service. ISE combines several traditions of evaluation and technology acceptance research and consists of five dimensions. In this study, we concentrate on dimension one, i.e. quality of information service, which includes perceived service as well as content quality evaluation and considers objective aspects (efficiency, effectiveness, gamification, functionality and usability) as well as users' perception (e.g. perceived usefulness and trust).

## 2.1 Heuristic usability evaluation

We applied a heuristic usability evaluation (Sarodnick & Brau, 2011: 144) to detect a VA's weaknesses relevant to the interaction with users. We determined ten relevant criteria covering the four objective categories in the ISE model (cf. table 1). We did not consider the aspect of gamification, as no VA offered any form thereof. After having contacted all three leading German developers, we were able to consult with Kauz Linguistic Technologies to ensure the design of a fitting criteria catalog, which could fit to the ISE model. To test the criteria, we designed queries corresponding to a VA's individual remit, i.e. VAs that specialize in sales were inquired about product consultation, VAs designed for information got questions regarding service. We determined the correctness of each reply given by the VA after having asked one query, with the exception for the categories "detail specification", "response calculation" and "context awareness". As the first two criteria require the VAs to interpret a combination of user information, two practice queries were asked first and a subsequent third one later, so that its reply by the VA could be evaluated. In terms of "context awareness", a dialogue with the VAs was established and their replies to the last query in that contextual dialogue was assessed.

For a service it is important that it is capable of "doing things right" (Schumann & Stock, 2014) to satisfy the customer or user. Schumann and Stock summarize relevant criteria in the category "efficiency". Aspects like correctness, swiftness of the disclosure and quality of information are relevant for an information service efficiency The VAs were tested for these qualities, especially in addition to the referral to possible useful and fitting sub-sites or pop-ups and the accurate usage of auto-correction.

Traditional effectiveness in retrieval research entails recall and precision studies (ibid.) – notably the last value considers a user's subjective view. In our user study (see chapter 2.2), we directly asked participants, if the VAs gave sufficient answers. The heuristic part considers more objective aspects important for successful VAs, which are autonomous information fishing (the VA's prompting for more and specific information from the user) and the interpretation of user intention (the VA's independent, flexible and accurate prediction and understanding of the user's input and needs).

As an additional aspect, Schumann and Stock depict the rating of the functionality as an "extent of its functions for information production and information searching", which make a service more valuable. Thus, we tested the VAs for further features, such as the recognition and processing of negation and response calculation. Response calculation means the VA's ability to combine different user input (e.g. the desired item, the cost of said item and the delivery options for that item) and provide appropriately accumulated information to the user, e.g.: User: "I'm looking for chocolate with fruits", VA: "Then I can offer you the Choconegro 1001 Nights, the Choconegro Creme Cointreau and the Chocoblanco Raspberry Brittle", user: "I prefer white chocolate", VA: "I'd like to offer you the Chocoblanco Raspberry Brittle", user: "I would like to buy white chocolate with fruits", VA: "Then I'd suggest you try the Chocoblanco Raspberry Brittle. It's our best white chocolate with fruits!". In addition, the VAs were tested for context awareness, meaning that they can associate referrals to preceding inputs, e.g.: "I like how the white chocolate looks." / "How much is it?" Here, "it" is referring to the white chocolate and the VA would still receive this input contextually and give a correct response.

Usability, i.e. serviceability of a VA, is the most important aspect for our study. In an ideal situation, the VA should be able to counteract any possible frustration on the part of its users. Therefore, the VA should identify a user's sentiment, allow for flexible wording of any input, recognize details in a user's inquiry and be aware of the context of each sentence or question.

We applied a scoring system to this evaluation. The highest achievable score for each criterion was set at 100 per cent, while the lowest most possible score resulted in 0 per cent. For each category we evaluated 50 VA replies to our queries. We considered two cases:

1. *Binary evaluation:* The examination of whether the reply of a VA was deemed correct in correspondence to the user input. It was only distinguished between 'yes' or 'no', or 'correct' and 'incorrect'.
2. *Gradation evaluation:* In some cases, several replies given by the VA contained correct and incorrect part. In this case, scales with a range of 0 (incorrect) to 2 (correct) were implemented, where an almost correct answer could be considered with the score of 1.

*Table 1: Assessment criteria and examples corresponding to the ISE categories*

| No. | Criteria | Examples | ISE-Category |
|---|---|---|---|
| 1. | Auto Correction | 'What options do I have?' → Understanding of the query despite the typo. | Efficiency |
| 2. | Forwarding | Being forwarded to the developers when specifically asked to be. | Efficiency |
| 3. | Autonomous Information Fishing | 'I need information.' → The VA proactively asks the user what they need information on. | Effectiveness |
| 4. | Interpretation of User Intention | 'I can't find any contact information.' → The VA provides the user with contact information instead of simply apologizing | Effectiveness |
| 5. | Sentiment Detection | 'I like you.' → The VA reacts positively. | Functionality |
| 6. | Negation | 'Do you offer shirts without a v-neck?' → The VA limits her offer to shirts without v-necks. | Functionality |
| 7. | Response Calculation | 'Do you offer vegan chocolate without nuts?' → The VA recognizes that the offered chocolate has to be vegan and without nuts at the same time. | Usability |
| 8. | Detail Specification | 'How long does shipping take within Germany?' → The VA delivers information on shipping. specifically in Germany. | Usability |
| 9. | Context Awareness | User: 'Do you offer a subscription?'<br>VA: 'Yes we do.'<br>User: 'How much would it cost?' → The VA recognizes that 'it' refers to 'subscription'. | Usability |
| 10. | Flexible Wording | 'When was your company founded' and 'Since when is your company active?' → The VA answers questions about a topic regardless of wording. | Usability |

### 2.2 Empirical user study

The user study took place in March and May 2016 and included search scenarios followed by a survey. Each participant evaluated one VA and had to pass two tasks (limited to eight minutes each). First, the participants had time to make themselves comfortable with the website and the VA. Afterwards, they were given a defined task (performing a sales conversation). In the second task, the participants were requested to ask the VA questions regarding service, products and the company the VA represented. After each task, the participants answered survey questions corresponding to the task. The survey included 50 questions (single and multiple choice, rating scales, free text) and was divided into six parts: Demographics, appraisal and expectation of VAs, experience during task one and two regarding information quality, general experience after tasks, and preferences on VA aesthetic. Additionally, the participants' voices and the mouse movements on the screen were recorded. In the following section, we concentrate on results regarding the four system quality aspects of the ISE model, which lead to the following survey questions using a 7-point Likert scale:

1. Simplicity/ease of use: The interaction with the virtual assistant felt … forced (1) – intuitive (7)
2a. Usefulness (questioned after 1st task): For my task, the information given by the VA were… useless (1) – useful (7)
2b. Usefulness (questioned after 2nd task): The information on products, services and the company was…useless (1) – useful (7)
3. Trust: I felt I was being taken seriously by the virtual assistant. I disagree (1) – I agree (7)
4. Fun: I had fun interacting with the virtual assistant. I disagree (1) – I agree (7)

## 3 Results

Table 2 shows the ten assessed criteria of the heuristic evaluation and the corresponding rate of correct uses by each VA in percentage. It has to be mentioned that the feature of forwarding to possibly helpful sub-sites was difficult to test because the quality of the sub-sites could not be assessed ap-

propriately regarding the question the VAs were tested on. However, all VAs are able to link to any sub-sites.

*Table 2: Assessed criteria of the heuristic evaluation*

| Category | Scores for Carla in % | Scores for Otto in % | Scores for Jana in % |
|---|---|---|---|
| 1. Auto Correction | 47.5 | 80 | 87.5 |
| 2. Forwarding | 100 | 100 | 100 |
| 3. Autonomous Information Fishing | 69.2 | 28.8 | 54 |
| 4. Interpretation of User Intention | 68.2 | 45.5 | 63.6 |
| 5. Sentiment Detection | 77.8 | 83.3 | 66.7 |
| 6. Flexible Wording | 62.1 | 55 | 53.8 |
| 7. Response Calculation | 100 | 25 | 25 |
| 8. Negation | 90 | 20 | 30 |
| 9. Detail Specification | 84.5 | 32.8 | 39.7 |
| 10. Context Awareness | 100 | 33.3 | 66.7 |

Regarding the user study, 44 students (23 female, 21 male, age 17 to 34) from diverse faculties (whereof 31 users had information science as part of their studies) participated. Carla and Jana were evaluated by 15, Otto by 14 participants. 9 of the 44 persons stated that they had experience with VAs, half of all participants said that the usage of a VA must be fun. This statement assumes that enjoyment is indeed a relevant influencing aspect, as Etemad-Sajadi (2014) already found out. After the interaction with the VAs, the participants were not quite satisfied: The medians of Otto and Jana is 3, only Carla has a median of 5 (fun, fig. 1).

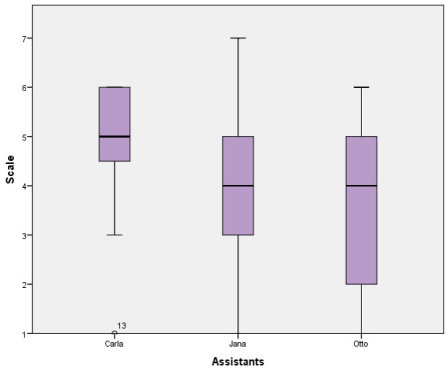


*Fig. 1* Boxplots of the survey questions on simplicity, trust and fun

The aspects of ease of use and trust are also shown in figure 1, which shows a boxplot comparison of the data recorded through the 7-point scales. Simplicity shows a higher scattering of Carla's data yet with a direction towards higher scores. Carla's median of 4 only slightly contrasts with the median of 3 from the Jana and 3.5 from Otto. Regarding trust, Carla's median of 5 contrasts with Jana's and Otto's median of 4.

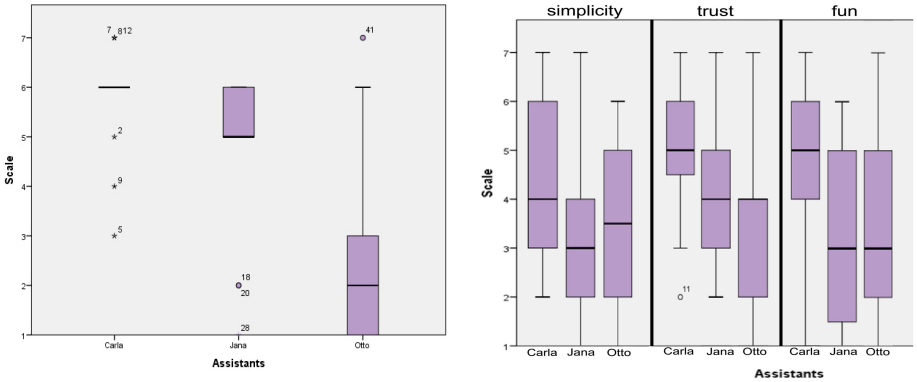


*Fig. 2* Usefulness after the 1st task (left) and after the 2nd task (right)

Concerning the perceived usefulness after task one (performing a sales conversation; fig. 2 left), Carla's delivered information was rated to be higher than the other two VAs, with a median of 6. However, participants' assessment thereof was less unanimous than for Jana or Otto. While Jana was rated similarly to Carla, with a median of 5, Otto's was rated worst with a median of 2. Regarding perceived usefulness after task two (open questions), the right diagram (fig. 2) displays a higher dispersion in the recorded data concerning the usefulness of Otto's information. Participants perceived both Jana and Otto as generally less useful as indicated by their medians of 4 in comparison to the information delivered by Carla, with a median of 5.

## 4 Discussion

Judging by Carla's replies to our queries, it is assumed that her superior natural language understanding is due to her linguistic capabilities that were designed and implemented differently from the other VAs'. Parts of the VAs'

processing techniques could be derived from questioning the VAs based on the ten criteria. It became obvious that Carla's strongest abilities were being able to identify the combination of information and keeping a conversational flow by seemingly being aware of what has already been mentioned in the interaction with the user. Additionally, Carla demonstrated her ability to deliver user tailored information and her understanding of negation in inquiries. While other aspects were deemed as good, Carla's difficulty with fittingly auto correct user questions became obvious. Otto appeared very capable of understanding user moods and looking past typos and other mistakes, yet often failed to understand a negation and the combination of information in a user's inquiry. Jana demonstrated her ability to autocorrect typos very efficiently and detects indirectly articulated intentions by the user. Yet she appeared to have difficulties with understanding negations and combined information. More specifically, regarding the criteria of response calculation, Carla analyzed every word in its grammatical form and its relation to other words in the sentence, meaning that she was programmed to process and interpret the morphological and syntactical significance of the user input. This resulted in a more precise handling of the user request. While some predetermined replies were observed, many answers seem to be given in real time. In contrast, Otto and Jana accommodate the ability to filter out keywords from the user input and match them with pre-existing articles that either functioned as the answer to a user inquiry or that were attached as an extra text section in addition to a generic message. It became apparent that due to the lack of syntactic analysis, recognizable keywords would still be processed, regardless if the user prepends or appends unintelligible strings of characters and digits. If two or more triggering keywords were entered, Otto and Jana addressed only one. The order in which these keywords were entered did not appear to affect which keyword was processed by the two VAs. It remains uncertain, whether an internal priority list for all known keywords exists or whether Otto and Jana favor specific keywords. The feature of auto correction was present in all VAs, while it appeared to have functioned more reliably with Otto and Jana, most likely due to their keyword-only analysis, since the lack of additional linguistic processing had no interfering effect on this feature. With autonomous information fishing, Carla proved to be the strongest with Otto being the weakest, as it rarely prompted the user for more information. Regarding interpretation of user intention, the three VAs seemingly featured a rather comparable understanding of the purpose of inquiries. Concerning sentiment detection, Otto proved to be the most aware of a user's

emotions, while Jana remained the most unaware. Furthermore, Carla proved the most flexible with wording of inquiries, although all three VAs allowed for an average flexibility. It was also obvious that Carla understands negation (e.g. "Do you offer vegan chocolate without nuts?"), detail specifications (e.g. "How long does shipping take within Germany?") and conversational context (e.g. anaphora resolution) the best, while Otto and Jana scored rather poorly in these categories.

The user study shows a similar trend concerning Carla's favorable position. Regarding the aspect of simplicity, figure 2 shows a higher scattering of Carla's data, yet with a direction towards higher scores, especially in comparison to Jana. The results of the participants' judgment of how seriously they were taken (trust), suggest Carla's distinct obligingness. Carla's median of five contrasts with Jana and Otto's median of four. Looking at the aspect of fun, interaction with Carla was deemed to be the most fun. In contrast, participants had slightly less fun with Otto and Jana, but their opinions were not unanimous, as the dispersion shows. Carla prevails in the assessment of usefulness, with Jana taking a close second place. Both VAs offer a wide palette of information on their offered services and goods. Otto's significantly worse score is most likely attributed to the fact that Otto does not deliver any information on the offered products and barely recognizes any product name. The usefulness examination based on task two displays a higher dispersion in Otto's data. Participants perceived both Otto and Jana as generally less useful in comparison to Carla, while simultaneously providing a rather unanimous assessment thereof. Both tasks reveal a discrepancy of the VAs' ability to offer useful information. Carla's scores are quite high, but not all participants agreed on this fact. The differences between both tasks may result from the diversity thereof. Some participants might not have felt familiar with the mission of the 1st task and might have been more confident in the 2nd task, where they could ask open questions to the VA. However, as the VA's usefulness is the most relevant aspect to use these services, the services should be improved to satisfy user needs.

In summary, Carla scored the best in both the heuristic as well as the user evaluation. She demonstrated a more natural understanding of the human language and offered a better developed understanding of users' information need overall. This was recorded in the qualitative user statements as well, such as: "I could hold a real conversation with Carla, and even after some seemingly conversation derailing remarks, she was able to pick up where we left off" (survey data, May 2016). Only the fact that she is designed to under-

stand inflection form of words interferes with her otherwise well working auto correction. An obvious weakness of Otto and Jana appears to be the response calculation and the recognition of negation in user input, which in turn was some of Carla's best features, as a user recalled: "The information was accurate and it was delivered quickly. The questions were understood by the assistant and I was positively surprised" (ibid.).

## 5 Conclusion and future work

We analyzed and compared three VAs concerning their usefulness and user perception. Therefore, we conducted a heuristic analysis as well as a user study. First results show that users generally perceived one VA more positively, which might be due to the VA's stronger language processing abilities. However, user opinions were quite heterogeneous. It should also be mentioned that with our tests the entirety of these VAs could not be tested due to the lack of insight into their natural language processing technologies, meaning that there is potential for future improved evaluation procedures. It will be valuable to compare these first outcomes with the recorded user voices and searches that will give detailed insights into user needs and their behavior during the interaction with a VA. In the future, we will analyze the users' perception of the VAs' aesthetic and its influence on enjoyment and trust.

**Acknowledgements**

We would like to thank Kauz Linguistic Technologies for their support with creating VA assessment categories and all participants of our user study.

## References

Ahamed, S. I., M. Sharmin, S. Ahmed, M. M. Haque, and A. J. Khan (2006): Design and implementation of a virtual assistant for healthcare professionals using pervasive computing technologies. In: *E & I Elektrotechnik und Informationstechnik*, 123 (4), 112–120. doi:10.1007/s00502-006-0335

Chao, C.-Y., T.-C. Chang, H.-C. Wu, Y.-S. Lin, and P.-C. Chen (2016): The interrelationship between intelligent agents' characteristics and users' intention in a search engine by making beliefs and perceived risks mediators. In: *Computers in Human Behavior*, 64, 117–125. doi:10.1016/j.chb.2016.06.031

Chang, J. W., M. C. Lee, and T. I. Wang (2016): Integrating a semantic-based retrieval agent into case-based reasoning systems: A case study of an online bookstore. In: *Computers in Industry*, 78, 29–42. doi:10.1016/j.compind.2015.10.007

Etemad-Sajadi, R. (2014): The influence of a virtual agent on web-users' desire to visit the company. In: *International Journal of Quality & Reliability Management*, 31 (4), 419–434. doi:10.1108/IJQRM-05-2013-0077

Fong, T., I. Nourbakhsh, and K. Dautenhahn (2003): A survey of socially interactive robots. In: *Robotics and Autonomous Systems*, 42 (3–4), 143–166.

García-Serrano, A. M., P. Martínez, and J. Z. Hernández (2004): Using AI techniques to support advanced interaction capabilities in a virtual assistant for e-commerce. In: *Expert Systems with Applications*, 26 (3), 413–426. doi:10.1016/j.eswa.2003.09.012

Keeling, K., P. McGoldrick, and S. Beatty (2010): Avatars as salespeople: Communication style, trust, and intentions. In: *Journal of Business Research*, 63 (8), 793–800. doi:10.1016/j.jbusres.2008.12.015

Kuligowska, K., and M. Lasek (2005): Intelligent agents in the human-computer interaction. In: B. Kubiak and A. Korowicki (Eds.): *Information management. Proceedings of the 7th International Conference on Information Management – Human Computer Interaction*. Gdańsk. https://www.chatbots.org/paper/intelligent_agents_in_the_human-computer_interaction/

Li, J. (2015): The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. In: *International Journal of Human-Computer Studies*, 77, 23–37. doi:10.1016/j.ijhcs.2015.01.001

McGoldrick, P. J., K. A. Keeling, and S. F. Beatty (2008): A typology of roles for avatars in online retailing. In: *Journal of Marketing Management*, 24 (3–4), 433–461. doi:10.1362/026725708X306176

Panetta, K. (2016): Gartner's Top 10 Strategic Technology Trends for 2017. http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/

Sarodnick, F., and H. Brau (2011): *Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung* (2nd Ed.). Bern: Huber.

Schumann, L., and W. G. Stock (2014): The information service evaluation (ISE) model. In: *Webology*, 11 (1).

Shiban, Y., I. Schelhorn, V. Jobst, A. Hörnlein, F. Puppe, P. Pauli, and A. Mühlberger (2015): The appearance effect: Influences of virtual agent features on performance and motivation. In: *Computers in Human Behavior*, 49, 5–11. doi:10.1016/j.chb.2015.01.077

Von der Pütten, A. M., N.C. Krämer, J. Gratch, and S.-H. Kang (2010): "It doesn't matter what you are!" Explaining social effects of agents and avatars. In: *Computers in Human Behavior*, 26 (6), 1641–1650. doi:10.1016/j.chb.2010.06.012